

Phenomena-Oriented Data Mining

Sara J. Graves (sgraves@cs.uah.edu)

Thomas H. Hinke (thinke@cs.uah.edu)

Heggere S. Ranganath (ranganat@cs.uah.edu)

John A. Rushing (jrushing@cs.uah.edu)

Information Technology and Systems Lab

The University of Alabama in Huntsville

Steven J. Goodman (steve.goodman@msfc.nasa.gov)

Paul J. Meyer (paul.meyer@msfc.nasa.gov)

Ravi Raghavan (ravi.raghavan@msfc.nasa.gov)

Franklin R. Robertson (pete.robertson@msfc.nasa.gov)

Global Hydrology and Climate Center

NASA / Marshall Space Flight Center

Objectives

- Development of a phenomena oriented data mining system.
- Generation of content-based metadata in addition to value-based metadata.
- Integration of content-based metadata with Information Management system.

Approach

- Mining for phenomena
 - » Class 1: Mining with known rules. When algorithms exist for identifying phenomena in a data set, those existing algorithms will be incorporated into the mining system.
 - » Class 2: Learning rules for recognizing phenomena based on samples. If no algorithm is known, but samples exist, pattern recognition techniques can be used to develop rules for identifying the phenomenon.
 - » Class 3: Identification of unknown phenomena. The mining system can be used to identify unusual events in the data that may not be associated with any known phenomenon.
- Loading metadata obtained from mining into spatial database

Class 1 Mining

- Mining with special purpose functions based on proven algorithms
- Mining with general purpose grid algebra operations
- Mining with general purpose window operations
- Mining with general purpose point operations

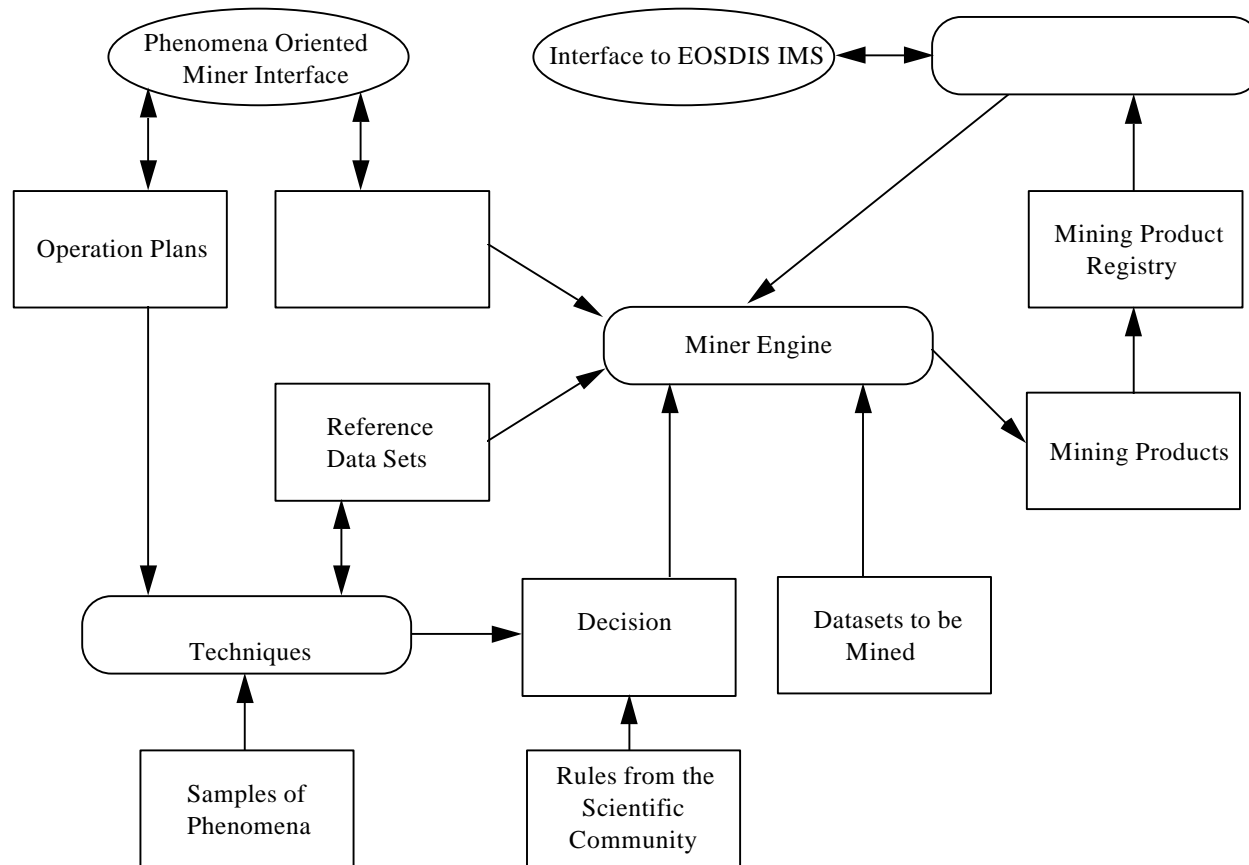
Class 2 Mining

- Extraction of features
- Development of decision rules through learning
- Implementation of decision rules
- Development of reference data sets

Class 3 Mining

- Identification of unusual events based on deviation from reference data
- Characterization of unusual events

Miner Architecture



Current Mining System Features

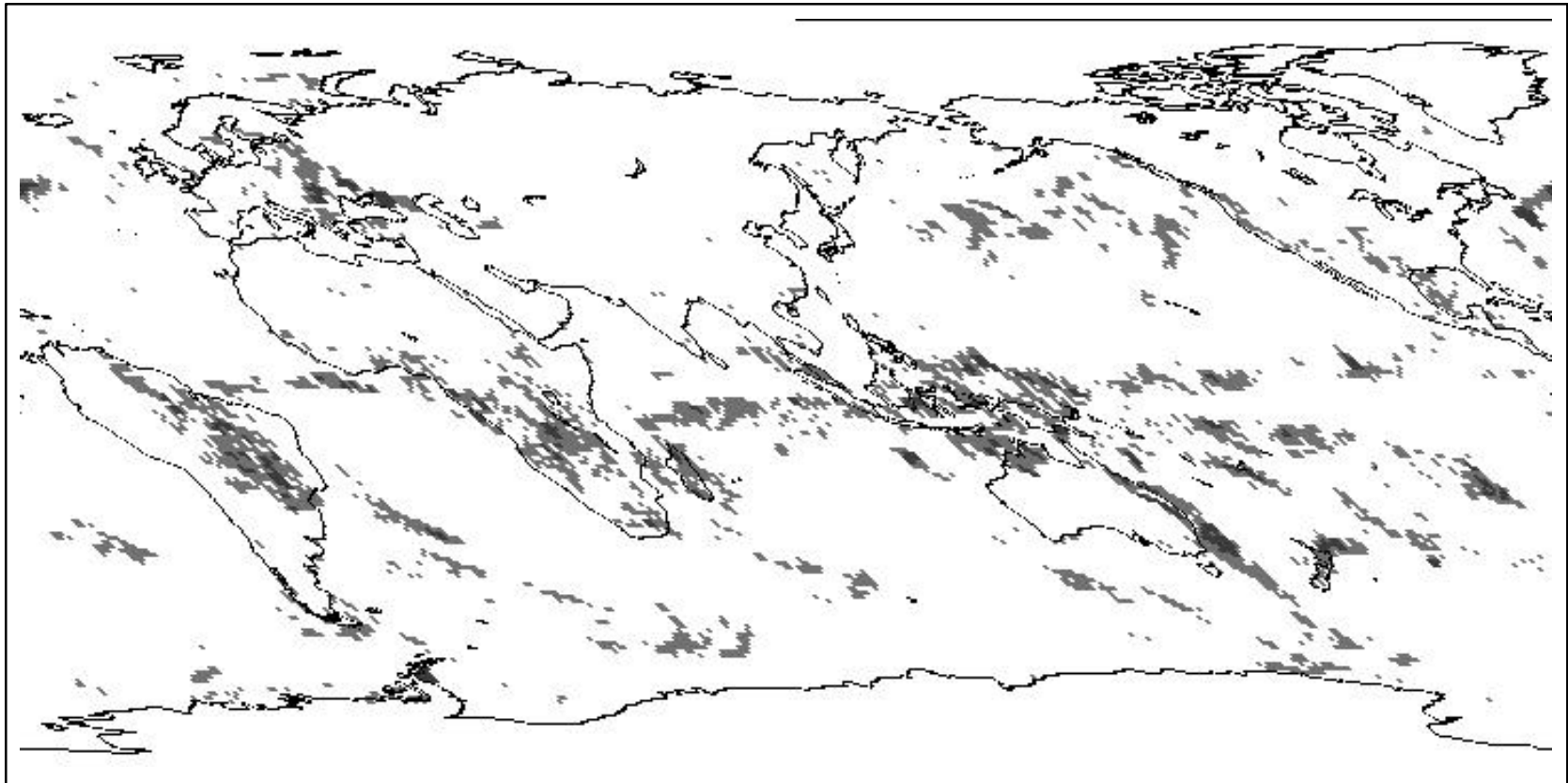
- Input Modules: PIP-2 CD, SSM/I Pathfinder, SSM/I TDR, US NEXRAD Rainfall
- Grid Algebra Operations: Select, Project, Aggregate
- Subsetting Capabilities: By space, time, data value
- Data Fusion Capabilities
- Scientific Algorithms: WetNet PIP SSM/I Rain Rate (Ferraro Algorithm), Mesoscale Weather System Detection (TAMU Algorithm)
- Clustering Algorithms: Maximin, K Means, Isodata
- Classifiers: Multiple Prototype Classifier
- Generation of phenomena enclosing polygons
- Ability to import polygons into spatial database
- WWW based user interface prototype

Mining Results

- Class 1 Mining Results
 - » Ferraro SSM/I rain rate data
 - » Mesoscale convective system data
- Class 2 Mining Results
 - » SSM/I Pathfinder reference data sets
 - » Multiple prototype classifier for oceanic rainfall in SSM/I data
- Metadata Search
 - » Polygons representing mesoscale convective systems loaded into DBMS
 - » Searches for phenomena of interest based on bounding boxes

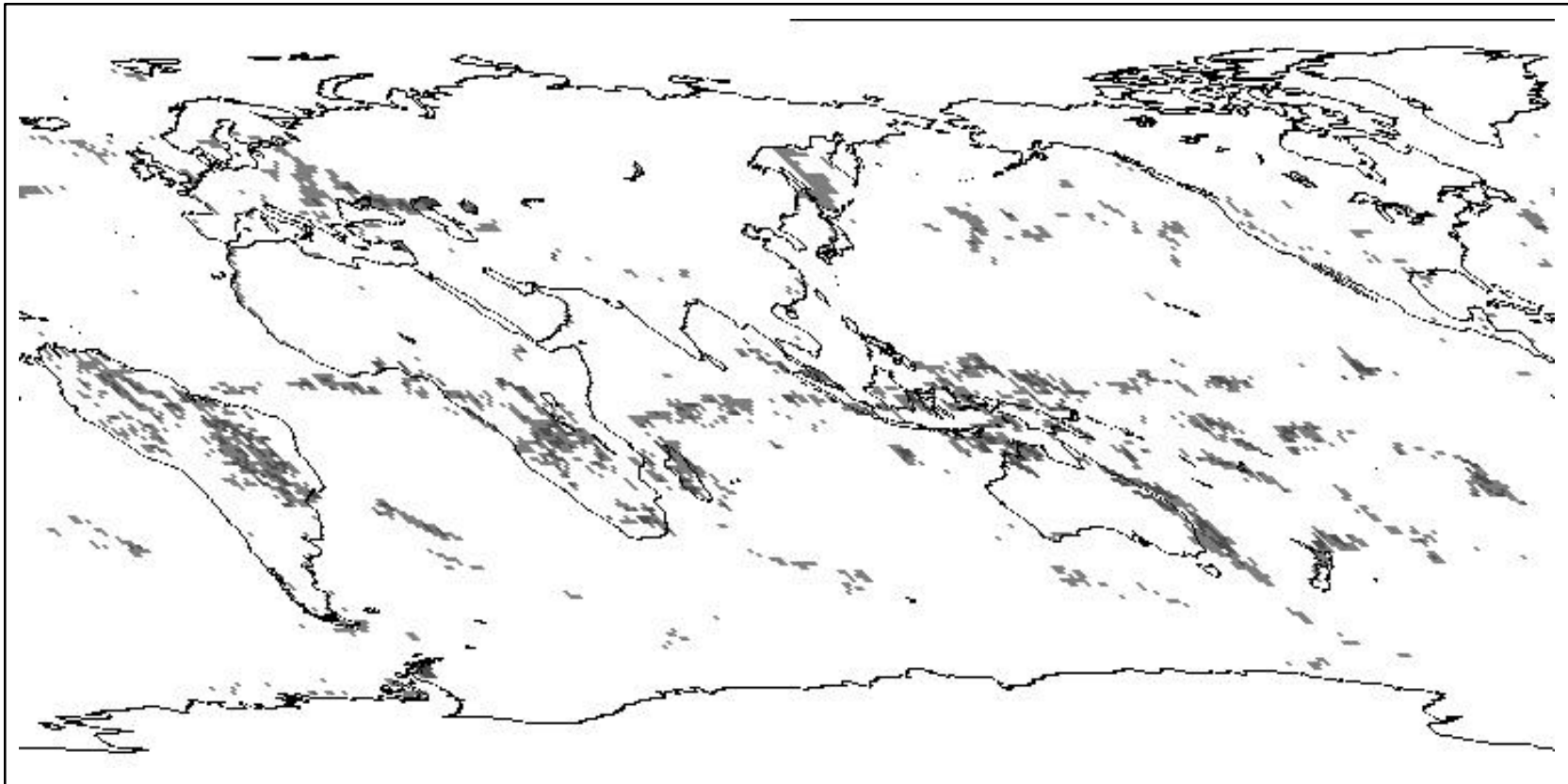
Mesoscale Convective Systems

February 7, 1988



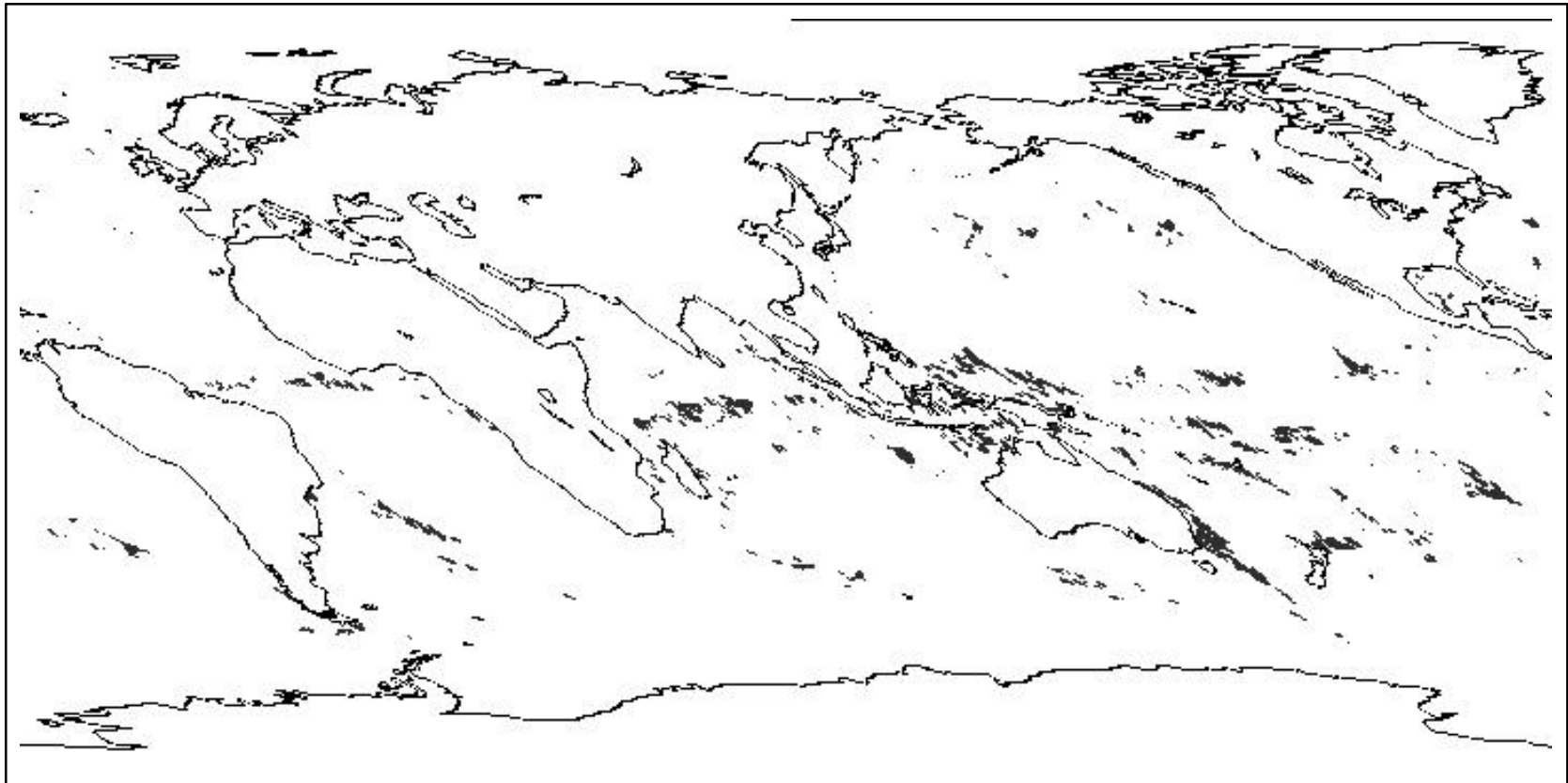
Ferraro Rain Rate

February 7, 1988



Oceanic Rain (Classifier)

February 7, 1988



Next Steps

- Enhance functionality by implementing a broader range of pattern recognition techniques
- Perform data fusion at the metadata level - multiple data sets, orbit models, external information
- Add capabilities to mine additional data sets and phenomena
- Combine various existing prototypes into an integrated data mining system
- Integrate with the IMS
- Begin development of initial parallel version of miner

ESDIS Support Desired

- Assist with identification of sample data for mining
- Assist, as needed, with interfaces to EOSDIS scientists
- Identification of appropriate involvement in telecons, meetings, and special interest groups
- Investigate ATM connectivity for prototype
- Mass storage enhancement